

Notice of Allowability

Application No.

10/675,756

Applicant(s)

NEVILL-MANNING ET AL.

Examiner

DIANE MIZRAHI

Art Unit

2165

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance (PTOL-85) or other appropriate communication will be mailed in due course. **THIS NOTICE OF ALLOWABILITY IS NOT A GRANT OF PATENT RIGHTS.** This application is subject to withdrawal from issue at the initiative of the Office or upon petition by the applicant. See 37 CFR 1.313 and MPEP 1308.

1. ☒ This communication is responsive to 10-17-08.
2. ☒ The allowed claim(s) is/are 4-6,8-12,15,19-23,25 and 26.
3. ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some* c) ☐ None of the:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. ____.
3. ☐ Copies of the certified copies of the priority documents have been received in this national stage application from the International Bureau (PCT Rule 17.2(a)).
- * Certified copies not received: ____.

Applicant has THREE MONTHS FROM THE "MAILING DATE" of this communication to file a reply complying with the requirements noted below. Failure to timely comply will result in ABANDONMENT of this application.

THIS THREE-MONTH PERIOD IS NOT EXTENDABLE.

4. ☐ A SUBSTITUTE OATH OR DECLARATION must be submitted. Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL PATENT APPLICATION (PTO-152) which gives reason(s) why the oath or declaration is deficient.
5. ☐ CORRECTED DRAWINGS (as "replacement sheets") must be submitted.
- (a) ☐ including changes required by the Notice of Draftsperson's Patent Drawing Review (PTO-948) attached
- 1) ☐ hereto or 2) ☐ to Paper No./Mail Date ____.
- (b) ☐ including changes required by the attached Examiner's Amendment / Comment or in the Office action of Paper No./Mail Date ____.
- Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the drawings in the front (not the back) of each sheet. Replacement sheet(s) should be labeled as such in the header according to 37 CFR 1.121(d).**
6. ☐ DEPOSIT OF and/or INFORMATION about the deposit of BIOLOGICAL MATERIAL must be submitted. Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Attachment(s)

1. ☐ Notice of References Cited (PTO-892)
2. ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3. ☒ Information Disclosure Statements (PTO/SB/08),
Paper No./Mail Date 6-24-08
4. ☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material
5. ☐ Notice of Informal Patent Application
6. ☒ Interview Summary (PTO-413),
Paper No./Mail Date 10-17-08.
7. ☒ Examiner's Amendment/Comment
8. ☒ Examiner's Statement of Reasons for Allowance
9. ☐ Other ____.

/Diane Mizrahi/
Primary Examiner, Art Unit 2165

DETAILED ACTION

This office action is in response to the newly submitted amendment filed October 15, 2008.

Examiner's Amendment

An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it **MUST** be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview with Attorney John Burbage on October 17, 2008.

SPECIFICATION AMENDMENT

Please amend paragraph [0034] as follows:

In one embodiment, the documents are then compared to create clusters of a number of closely related documents. For example, each sample document may be compared and then aligned with the closest nine documents to create a cluster of ten documents. In one embodiment, the dynamic programming alignment algorithm is used to compare the documents. The dynamic programming alignment algorithm compares and aligns documents to compute relative scores for the compared documents

known as edit distances. The edit distance can be a number, roughly proportional to the number of insertions and deletions necessary to transform one document into another. An alignment of two documents, for example, can be a list of those insertions or deletions, or equivalently, a mapping from parts of one document to parts of another. Dynamic programming alignment is a method understood by those skilled in the art, and accordingly need not be described in further detail herein. ~~Further details on dynamic programming may be found in Dan Gusfield, Algorithms on Strings, Trees, and Sequences, Computer Science and Computation Biology (Cambridge University Press 1997), which is incorporated herein by reference.~~

CLAIM AMENDMENTS

1. (Canceled).
2. (Canceled).
3. (Canceled).
4. (Previously Presented) The method of claim 12, further comprising determining a cluster of related articles from the related articles.
5. (Previously Presented) The method of claim 4, wherein determining the cluster of related articles is performed by:
 - using the dynamic programming alignment algorithm to compute edit distances between the seed article and the related articles; and
 - choosing the cluster of related articles based on the edit distances.
6. (Original) The method of claim 4, wherein the identifying at least one information field within the seed article is performed by comparing the seed article to the cluster of articles.
7. (Canceled)

8. (Previously Presented) The method of claim 12, wherein the articles are web pages.

9. (Original) The method of claim 8, wherein the related articles are web pages on a web site.

10. (Original) The method of claim 9, further comprising simplifying the content on a web page.

11. (Original) The method of claim 10, wherein simplifying the content includes preserving visible text, visible images, and visible paragraph and table formatting.

12. (Currently Amended) A method for information extraction, comprising:

- accessing a plurality of related articles;
- determining a seed article from the related articles, the seed article containing variable data;
- identifying at least one information field within the seed article by comparing the seed article to at least one other related article, the comparison comprising using a dynamic programming alignment algorithm to determine an alignment between the seed article and the related article;
- creating a template based on the identified information field, the template identifying an information field in the related articles corresponding to the variable data;

identifying a plurality of templates each comprising at least one information field;

comparing a source article to the templates to determine a closest template;

associating data from the source article with an information field from the closest template; and

extracting the associated data.

13. - 14. (Canceled).

15. (Currently Amended) A method of extracting data from a source article, comprising:

identifying a plurality of templates each comprising at least one information field corresponding to variable data in articles;

comparing the source article to the templates to determine a closest template, wherein comparing the source article to the templates is performed by a dynamic programming alignment algorithm to compute an edit distance between the source article and the templates;

associating data from the source article with an information field corresponding to variable data from the closest template;

extracting the associated data; and

displaying the associated data.

16. - 18. (Canceled)

19. (Previously Presented) The computer program product of claim 23, wherein comparing the seed article to at least one other related article is performed by using the dynamic programming alignment algorithm to determine an alignment between the seed article and the related article.

20. (Previously Presented) The computer program product of claim 23, further comprising computer program code for determining a cluster of related articles from the related articles.

21. (Previously Presented) The computer program product of claim 20, wherein determining a cluster of related articles is performed by:

using the dynamic programming alignment algorithm to compute edit
distances between the seed article and the related articles; and
choosing the cluster of related articles based on the edit distances.

22. (Previously Presented) The computer program product of claim 20, wherein the identifying at least one information field within the seed article is performed by comparing the seed article to the cluster of related articles.

23. (Currently Amended) A computer program product having a tangible computer-readable storage medium having ~~computer-executable~~ processor-executable code encoded thereon for performing information extraction when executed by a processor, the ~~computer-executable~~ processor-executable code comprising code for:

accessing a plurality of related articles;

determining a seed article from the related articles, the seed article containing variable data;

identifying at least one information field within the seed article by comparing the seed article to at least one other related article;

creating a template based on the identified information field, the template identifying an information field in the related articles corresponding to the variable data;

identifying a plurality of templates each comprising at least one information field;

comparing a source article to the templates to determine a closest template, the comparison comprising using a dynamic programming alignment algorithm to compute an edit distance between the source article and the templates;

associating data from the source article with an information field from the closest template; and

extracting the associated data.

24. (Canceled).

25. (Previously Presented) The computer program product of claim 23, further comprising code for:

displaying the associated data.

26. (Previously Presented) The computer program product of claim 23, further comprising code for:

storing the associated data.

Based on Applicant's arguments, Examiner formally withdraws the objection to the specification.

Allowable Subject Matter

Claims 4-6, 8-12, 15, 19-23, and 25-26 are allowed over the prior art made of record.

The following is an examiner's statement of reasons for allowance:

Regarding Independent Claims 12, 15 and 23, under the broadest reasonable interpretation of the claimed limitation consistence with the Applicant's Specification, the prior art cited in the record fails to teach all of the Applicant's claimed limitation. In particular, the claimed invention advantageously provides a finer level of detail that enables for determining a seed article containing variable data from the related articles, identifying an information field within the seed article by comparing the seed article to the other related article, using a dynamic programming alignment algorithm to determine an alignment between the seed article and the related article, creating a template based on the identified information field corresponding to the variable data,

Art Unit: 2165

identifying templates, and comparing a source article to the templates to determine a closest template.

Thus, prior art of record neither render obvious nor anticipates the combination of claimed elements in light of the specification.

Dependent claims 4-6, 8-11, 19-22, and 25-26 are allowed at least by virtue of their dependencies from their pertinent independent claims.

After a further search and a thorough examination of the present application and in light of the prior art made of record, claims 4-6, 8-12, 15, 19-23, and 25-26 are allowed.

Any comments considered necessary by applicant must be submitted no later than the payment of the issue fee and, to avoid processing delays, should preferably accompany the issue fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for Allowance."

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Diane D. Mizrahi whose telephone number is 571-272-4079. The examiner can normally be reached on Monday-Thursday (9:30 - 4:30 p.m.).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Christian Chase can be reached on (571) 272-4190. The fax phone numbers for the organization where this application or proceeding is assigned are (703) 872-9306 for regular communications and (703) 305-3900 for After Final communication.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (571) 272-2100.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll free).

/Diane Mizrahi/

Diane.Mizrahi@USPTO.gov
Primary Patent Examiner
Technology Center 2100

Application/Control Number: 10/675,756
Art Unit: 2165

Page 12

October 17, 2008